Structural Studies of Substrate- and Product-Complexes of 5-Aminolaevulinic Acid Dehydratase from Humans, *E. Coli* and the Hyperthermophile *Pyrobaculum Calidifontis*.

N. Mills-Davies,[1] D. Butler,[1] E. Norton,[1] D. Thompson,[1] M. Sarwar,[1] J. Guo,[2] R. Gill,[2] N. Azim,[3] A. Coker,[2] S. P. Wood,[2] P. T. Erskine,[2,4] L. Coates,[5] J. B. Cooper,[2,4*] N. Rashid,[3] M. Akhtar[3] and P. M. Shoolingin-Jordan.[1]

[1] *School of Biological Sciences, University of Southampton, Southampton, SO16 1BJ, UK.*

[2] *Laboratory of Protein Crystallography, Drug Discovery Group, Wolfson Institute for Biomedical Research, UCL Division of Medicine, London, WC1E 6BT, UK.*

[3] *School of Biological Sciences, University of the Punjab, Lahore-54590, Pakistan.*

[4] *Department of Biological Sciences, Birkbeck, University of London, Malet Street, London, WC1E 7HX, UK.*

[5] *Biology and Soft Matter Division, Oak Ridge National Laboratory, Oak Ridge, TN 37831, USA.*

* For correspondence (e-mail: jon.cooper@ucl.ac.uk).

Corresponding author details

Professor Jonathan B. Cooper, Laboratory for Protein Crystallography, Drug Discovery Group, Wolfson Institute for Biomedical Research, UCL Division of Medicine, London, WC1E 6BT, UK and Department of Biological Sciences, Birkbeck, University of London, Malet Street, London, WC1E 7HX, UK.

Tel: 07883 861983,

E-mail: jon.cooper@ucl.ac.uk,

Web: https://www.ucl.ac.uk/wibr/research/drug-discovery

Keywords

Tetrapyrrole biosynthesis, chlorophyll, haem, 5-aminolaevulinic acid dehydratase, porphobilinogen synthase, protein crystallisation, X-ray structure.

Synopsis

We report X-ray structures of the enzyme 5-aminolaevulinate dehydratase (ALAD) purified from human blood, as well as the recombinant human form and the ALAD from the hyperthermophilic archaeon *Pyrobaculum calidifontis*. In addition, the structure of *E. coli* ALAD co-crystallised with a non-covalently bound moiety of the product, porphobilinogen (PBG) has been determined. The structures give insight into the zinc-dependence of these enzymes, the possible role of electrostatics in substrate funnelling and details of the interactions made by the bound pyrrole end-product.

Abstract

We report a number of X-ray analyses of an enzyme involved in a key early stage of tetrapyrrole biosynthesis. Two structures of human 5-aminolaevulinate dehydratase (ALAD), native and recombinant, have been determined at 2.8 Å resolution, showing that the enzyme adopts an octameric quaternary structure in accord with previously published analyses of the enzyme from a range of other species. However, this is in contrast to the finding that a disease-related F12L mutant of the human enzyme uniquely forms hexamers (Breinig, S. *et al*., (2003) *Nat. Struct. Biol*. **10**, 757-763). Monomers of all ALADs adopt the TIM barrel fold; the subunit conformation that assembles into the octamer includes the N-terminal tail of one monomer curled around the $(\alpha/\beta)_8$ barrel of a neighbouring monomer. Both crystal forms of the human enzyme possess two monomers per asymmetric unit, termed A and B. In the native enzyme, there are a number of distinct structural differences between the A and B monomers, with the latter exhibiting greater disorder in a number of loop regions and in the active site. In contrast, the second monomer of the recombinant enzyme appears to be better defined and the active site of both monomers clearly possesses a zinc ion which is bound by three conserved cysteine residues. In native human ALAD, the A monomer also has a ligand resembling the substrate ALA which is covalently bound by a Schiff base to one of the active site lysines (Lys 252) and is held in place by an ordered active site loop. In contrast, these features of the active site structure are disordered or absent in the B-subunit of the native human enzyme. We also report on the octameric structure of the zinc-dependent ALAD from the hyperthermophile *Pyrobaculum calidifontis* (Pc-ALAD) at a somewhat lower resolution of 3.5 Å. Finally, we present the details of a high resolution structure of the *E. coli* ALAD enzyme co-crystallised with a non-covalently bound moiety of the product, porphobilinogen (PBG). This structure reveals that the pyrrole side chain amino group is datively bound to the active site zinc ion and that the PBG

carboxylates interact with the enzyme via hydrogen bonds and salt-bridges with invariant residues. A number of hydrogen bond interactions that were previously observed in the structure of yeast ALAD with a cyclic intermediate resembling the product PBG appear to be weaker in the new structure suggesting that these interactions are only optimal in the transition state.

1. Introduction

Most haem in living systems is derived from endogenous synthesis, with each tissue producing it according to its own needs. In humans, the bone marrow accounts for more than 70% of total haem synthesis and is the site of haem incorporation into haemoglobin for production of erythrocytes (Ajioka *et al*., 2006). The liver is the second most important site, accounting for approximately 15% of the total haem synthesised. In this organ there is a high requirement for haem due to the many mitochondrial cytochromes and other haemoproteins, such as cytochrome P450, catalase and cytochrome b_5.

All of the enzymes catalysing the eight steps of the haem biosynthesis pathway have now been analysed structurally (for review see Layer *et al.*, 2010). One of the first enzymes to be studied was 5-aminolaevulinic acid dehydratase (E.C.4.2.1.24) which is also referred to as porphobilinogen synthase since it catalyses the condensation of two 5-aminolaevulinic acid (ALA) molecules to form the pyrrole porphobilinogen (PBG) (Fig. 1) (Jordan, 1991; 1994; Warren & Scott, 1990; Jaffe, 1995; Jaffe & Lawrence, 2014; Jaffe, 2016). Four molecules of the product of ALAD (PBG) are condensed in a reaction catalysed by porphobilinogen deaminase to form the linear tetrapyrrole, preuroporphyrinogen. This is then cyclised and rearranged by uroporphyrinogen synthase to give uroporphyrinogen III, the first macrocyclic tetrapyrrole in the pathway. These three steps from ALA to uroporphyrinogen III are common to the biosynthesis of haem, chlorophyll, cobalamins and all other tetrapyrroles.

Single-turnover experiments on ALAD have shown that the first substrate molecule to bind to the enzyme ultimately forms the ‘propionate’ half of the product PBG, whilst the second substrate molecule forms the ‘acetate’ half of PBG (Jordan & Gibbs, 1985). This has led to the widely used terminology of the ‘P’ and ‘A’ binding sites in the enzyme which bind the substrate moieties forming the propionate and acetate halves of the product, respectively (Fig. 1). There is evidence that many ALADs require a metal ion for activity with animal enzymes using zinc and plant enzymes having at least a partial requirement for magnesium (Jaffe, 2003). Representatives of both classes exist in bacteria.

The gene for human ALAD is situated on chromosome 9q33.1 (Wetmur *et al*., 1986; Potluri *et al*., 1987) and contains two alternative 5' non-coding exons, 1A and 1B, which are followed by 11 coding exons, numbered 2 to 12 (Kaya *et al*., 1994). Exon 1A is incorporated into the housekeeping transcript whereas 1B is incorporated into the erythroid-specific transcript, although the proteins produced from both of these transcripts are identical. Hereditary deficiencies in ALAD are associated with the rare genetic disease Doss or ALAD porphyria (Doss *et al*., 1979). This is an autosomal recessive disorder that has severe neurological symptoms due to the accumulation of substrate 5-aminolaevulinate which structurally resembles the neurotransmitter γ-aminobutyric acid (Sassa, 1998). The zinc ions in ALAD are displaced upon addition of lead, a potent inhibitor of the zinc-dependent enzymes (Jaffe *et al*., 2001; Simons, 1995; Warren *et al*., 1998). Inhibition of ALAD by lead ions is one of the major manifestations of acute lead poisoning which often leads to neurological and psychotic disturbances. The enzyme is also inhibited by succinylacetone which accumulates in patients with type-I tyrosinaemia (Lindblad *et al*., 1977) - approximately 40 % of children with this disease develop porphyric symptoms. Relative to other enzymes in the haem biosynthesis pathway, ALAD is present in large excess e.g. the activity of ALAD in liver is 80 – 100 times that of the preceding enzyme in the pathway. These effects may account for the fact that heterozygotes for ALAD deficiency are asymptomatic but are likely to be more prone to environmental or occupational lead poisoning. At least 11 mutations causing ALAD porphyria have been identified and a significant proportion of sufferers are compound-heterozygotes i.e. each ALAD allele has a different harmful mutation. Porphyria-associated mutants of ALAD have a tendency to adopt the hexameric state (Jaffe & Stith, 2007). There have been several reports that ALAD co-purifies with the proteasome and that it affects the activity of these large protein complexes which catalyse ATP-dependent proteolysis and play key roles in cell-cycle control and homeostasis (Guo *et al*., 1994; Bardag-Gorse & French, 2011). An interaction with the chaperone Hsp70 has also been reported (Gross *et al*., 1999).

All ALADs typically contain around 330 residues per subunit and are usually octameric (Jordan, 1991; 1994; Warren & Scott, 1990; Jaffe, 1995; 2016). Studies on bovine and human ALAD showed that catalysis occurs through the formation of a Schiff base link at the P-site between the 4-keto-group of substrate and an invariant lysine residue, Lys 252 in human ALAD (Gibbs & Jordan, 1986). Later structural studies established the importance of a second invariant lysine residue (Lys 199 in human ALAD) in forming a Schiff base with the A-side substrate (Erskine *et al*., 1997a; 2001a). The implications of these structural studies for the mechanism have been reviewed critically by Jaffe

(2004; 2016) as well as by Erdtman *et al*., (2010) and Tian *et al*., (2012) who used theoretical approaches to investigate the order of bond formation during the reaction.

The human and bovine ALAD enzymes have been purified in high yield from both blood and liver where the enzyme is present in high abundance (Gibbs *et al*., 1985; Jordan & Seehra, 1986). Gel-filtration, sedimentation, electron microscopy and solution scattering studies have shown that the mammalian ALAD enzymes are octameric (Wu *et al*., 1974; Bevan *et al*., 1980; Pilz *et al*., 1988). However, there are reports that the plant enzyme can adopt a hexameric form (e.g. Kokona *et al*., 2008) and a hexameric, porphyria-associated mutant form of the human enzyme has been described (Breinig *et al*., 2003) leading to speculation that the inter-conversion between these oligomeric states provides a means of regulating the activity of ALAD enzymes, as corroborated by further studies (Jaffe & Lawrence, 2012).

Some zinc-dependent ALADs can bind multiple metal ions per subunit (Jaffe *et al*., 1995; Spencer & Jordan, 1994). One of the metal ions is essential for catalytic activity and another can have an allosteric activating role. The catalytic metal-binding site is now known to consist of three conserved zinc-binding cysteine residues. With the exception of animals and fungi, most zinc-dependent ALADs also contain an allosteric magnesium site. In contrast, ALADs from plants, apicoplasts, and some prokaryotes have the allosteric magnesium site, but lack several of the active site cysteines. Some of these ALADs are catalytically competent in the absence of active site metal ions, but the activity is dramatically stimulated by addition of magnesium.

High resolution X-ray structures have been determined for ALADs from yeast (Erskine *et al*., 1997a), *E. coli* (Erskine *et al*., 1999), *P. aeruginosa* (Frankenberg *et al*., 1999), the green sulphur bacterium *Chlorobium vibrioforme* (Coates *et al*., 2004) and the protozoan parasite *Toxoplasma gondii* (Jaffe *et al*., 2011). The latter enzyme possesses extensions at the N- and C-terminal ends which form elements of secondary structure which were not seen in earlier structures but are clearly involved in quaternary interactions. It also provided one of the clearest pictures of the product porphobilinogen bound to the enzyme active site. In addition, structures are available for mouse ALAD (PDBIDs: 2Z0I and 2Z1B), the human erythrocyte enzyme (PDBID: 1E51) and a porphyria-associated mutant of human ALAD which forms hexamers with reduced catalytic activity (Breinig *et al*., 2003). A large number of ALAD inhibitors have been structurally characterised bound to the enzyme from various species including the antibiotic alaremycin which is produced by *Streptomyces* and has been shown to operate by inhibiting bacterial haem synthesis (Heinemann *et al*., 2010). All of these inhibitors act by forming Schiff base complexes with one or both of the active site lysines (Cooper & Erskine, 2004). Recently,

it has been suggested that a number of heterocyclic inhibitors of ALAD have potential in treatment of infections by filial nematodes – parasitic organisms which are unable to synthesise tetrapyrroles and therefore rely on the endosymbiotic α-proteobacterium *Wolbachia* for haem synthesis (Lentz *et al*., 2013; 2014).

P. calidifontis strain VA1 is a hyperthermophilic archaeon belonging to the genus *Pyrobaculum* (Amo *et al*., 2002). Whilst this genus has a large diversity of oxygen requirements ranging from strictly anaerobic to obligate aerobic, this particular species is known to be a facultative anaerobe. Interest in the oxygen needs of this organism led us to study enzymes responsible for the synthesis of heme which is central to oxygen assimilation.

In this paper we describe two previously unpublished structure determinations of human erythrocyte ALAD using protein purified from blood and expressed in *E. coli*. We also describe the structure analysis of the *P. calidifontis* enzyme and a high resolution structure of *E. coli* ALAD co-crystallised with a non-covalently bound moiety of the product, porphobilinogen.

2. Methods and materials

2.1 Cloning and expression

For expression of recombinant human ALAD, a synthetic gene of 990 base pairs, encoding the 330 amino acids of human ALAD was constructed from a total of 18 oligonucleotides, 9 primers for each strand. The primers had an overlap of approximately 10 base pairs. Only the codon specifying Pro 3 was changed, from CCC to CCG, as CCC is known to be a weakly expressing Pro codon in *E. coli*. The 5’ end of the gene was engineered with a *Nde*I restriction site and the 3’ end with a *Bam*HI restriction site. These restriction sites enabled insertion of the cDNA into the vector pT7-7 which was then transformed into the bacterial strain, BL21 (DE3) and gene expression was induced by induction with 1 mM isopropyl-β-D-thiogalactopyranoside (IPTG).

The putative gene sequence for Pc-ALAD was obtained from the KEGG database and was used to design primers for amplification of the *hemB* gene. The forward primer, 5' **CATATG**CGCGTACAGTTTCCAACGAC 3', was designed from the N-terminal sequence of the gene with the introduction of a cleavage site for the restriction enzyme *Nde*1 (shown in bold) at the 5' end of the primer. The reverse primer, 5' GCGCAAGGAGTAGCGCTGTTCG 3' was designed using the C-terminal end of the gene sequence and did not contain any restriction sites. The gene was amplified using standard PCR methods, ligated into the vector pTZ57R/T by TA cloning (Holton &

Graham, 1991) and its orientation with respect to the polycloning site was confirmed by double restriction digestion and sequencing. The gene was then excised from this plasmid using *Nde*I and *Bam*HI for ligation into the expression vector pET21d which had previously been cleaved with the same two restriction enzymes. The gene was expressed using *E. coli* BL21-CodonPlus cells using standard methods with addition of the inducer IPTG to a final concentration of 0.2 mM at mid-log phase followed by growth of the cultures for a further 4 hours.

Expression and purification of endogenous *E. coli* ALAD has been described previously (Spencer & Jordan, 1993).

2.2 Purification

Native human ALAD was purified from erythrocytes using the method of Gibbs *et al*., (1985) and the assay described by Mauzerall & Granick (1956). Initially 1.2 litres of buffy coat residues were washed in isotonic saline containing 1 mM β-mercaptoethanol (BME) followed by centrifugation at 5,000 rpm. The cells were then lysed in de-ionised water containing 40 mM BME and the solution was sonicated for 60 seconds at medium power in an ice/water mixture prior to removal of the cell debris by ultracentrifugation at 48,000 g. The supernatant was collected and a stock solution of 1 M potassium phosphate pH 7.5 was added to give a final buffer concentration of 10 mM. This solution was then mixed for 1.5 hours with DEAE-Biogel A ion-exchange resin that had been pre-equilibrated with the same buffer. The resin was collected by centrifugation and washed three times with the buffer. Weakly bound proteins were then removed with 50 mM KCl in the same buffer and the more strongly bound ALAD was eluted by addition of 350 mM KCl. Raising the ammonium sulphate (AS) concentration to 60 % saturation allowed the enzyme to be precipitated and separated from impurities (mainly haemoglobin) by centrifugation at 12,000 g. A buffer solution of 100 mM potassium phosphate at pH 6.8 and 20 mM BME was used to re-suspend the pellet of ALAD giving a final protein concentration of 30 mg/ml. Further removal of impurities involved heat-treating the enzyme at 60 °C for 3 minutes with stirring followed by rapid cooling to below 10 °C with an ice-water mix and then centrifuging at 12,000 g for 10 minutes. The supernatant was removed and AS was added to 30 % saturation to precipitate ALAD which was collected by centrifugation and re-suspended in 30 mM phosphate buffer at pH 7.9 and dialysed overnight against 20 litres of the same buffer. The dialysed enzyme was applied to a DEAE-Biogel A column and eluted by application of a 0 – 0.4 M KCl gradient in 30 mM phosphate buffer at pH 7.9. Active fractions were precipitated by addition of AS to 60 % saturation and collected by ultracentrifugation at 48,000 g prior to re-suspension in 50 mM phosphate buffer pH 6.8 and a second dialysis step against the same buffer. The protein was then loaded onto a Biorad HTP hydroxyapatite column, previously equilibrated with 50 mM potassium

phosphate at pH 6.8 and eluted by applying a gradient of 50 – 500 mM phosphate buffer. Active fractions were collected and concentrated exactly as before prior to re-suspension in 60 mM phosphate buffer pH 6.8 and 0.5 M KCl. A final gel-filtration step was conducted using a Biogel A-0.5M column prior to storage of the AS-pelleted enzyme at 4 °C in a sealed tube under nitrogen gas. All buffers used in the purification included BME at a concentration of 25 mM.

Purification of recombinant human ALAD from the *E. coli* cell supernatant involved fractional AS precipitation with impurities being removed by addition of AS to 35 % saturation and the protein of interest precipitated by addition of AS to 60 % saturation. After dialysis, the partially purified enzyme was applied to a Bio-Rad hydroxyapatite column and eluted with a linear gradient of 50 – 300 mM potassium phosphate pH 6.8, containing 5 mM BME, 50 μM $ZnCl_2$ and 100 μM PMSF. The active fractions were pooled, AS-precipitated and finally applied to a Bio-Gel A-0.5M gel-filtration column.

Purification of the recombinant Pc-ALAD from the *E. coli* cell supernatant involved heat treatment to a temperature of 80 °C for 20 minutes followed by ultracentrifugation at 40,000 g for 15 minutes to remove impurities. The Pc-ALAD was then precipitated by increasing the AS concentration to approximately 35 % saturation followed by ultracentrifugation as above. The enzyme was then dissolved in 50 mM potassium phosphate buffer pH 7.9 containing 5 mM BME and 100 μM $ZnCl_2$. The dissolved protein was dialyzed against the same buffer to remove all traces of residual AS. The protein was then subjected to anion exchange using a HiTrapQ FF column by applying a 0 – 1 M NaCl gradient in the above buffer solution and the Pc-ALAD fractions, which eluted at around 0.4 M NaCl, were stored at 4 ˚C after dialysis to remove the NaCl.

The zinc ion content of Pc-ALAD was determined by atomic-absorption spectrophotometry at a wavelength of 213.8 nm using an air-acetylene flame. A standard curve for zinc ion concentrations in the range 0.00 - 0.03 mM was prepared with known amounts of $ZnCl_2$ dissolved in deionized water. Pc-ALAD samples were prepared by extensive dialysis against buffers devoid of zinc ions and a protein blank was also made by precipitation of a 1 mg/ml solution of Pc-ALAD with trichloroacetic acid in order to remove any bound zinc ions and the pellet was washed before being dissolved in 8 M urea. All samples and buffers were passed through a syringe filter with 0.22 μm pore size before analysis.

2.3 Crystallisation

For crystallisation of both native and recombinant human ALAD, the enzyme was equilibrated with 10 mM Tris pH 7.4 containing 10 mM DTT and 100 μM $ZnCl_2$ by dialysis and then concentrated by

centrifugal ultra-filtration. Crystals were grown by the hanging-drop method with the protein at a concentration of 10 mg/ml mixed in 50:50 ratio with a well solution consisting of 0.1 M MES buffer in the pH range 6.2 – 6.5, 1 – 1.6 M AS and 0 - 10 % dioxane. The resulting crystals were cryo-protected by addition of glycerol to approximately 30 % v/v and were frozen in a liquid ethane bath in liquid nitrogen.

Screening for crystallisation conditions for Pc-ALAD using the hanging-drop method involved replacing the buffer with 10 mM Tris-HCl pH 8.0 using a Vivaspin centrifugal concentrator. At a concentration of 5 mg/ml, the protein crystallised in the presence of 1 mM levulinic acid with 0.1 M HEPES pH 7.5, 10 % isopropanol and 20 % polyethylene glycol (PEG) 4000 in the well solution. Crystals were cryo-protected by immersion in a mixture of paraffin and Paratone^{TM} oil prior to mounting in loops and were flash-cooled using an Oxford Cryosystems cryostream.

The crystallisation of native *E. coli* ALAD has been reported in detail previously (Erskine *et al*., 1997b). Co-crystals of the enzyme with PBG were obtained using the hanging drop method by mixing the protein at a concentration of 7 mg/ml with an equal volume of well solution made of 200 mM Tris pH 8.0 – 8.4, 2 % saturated AS, 200 μM zinc sulphate, 6 mM BME and 3 mM PBG. The co-crystals were cryo-protected with 30 % glycerol and flash-cooled in liquid ethane prior to storage under liquid nitrogen.

2.4 Data collection and structure determination

Data collection on the crystals of native human ALAD at a temperature of 100 K was undertaken at the SRS (Daresbury, UK) using beam line 9.5 equipped with a Marresearch 345 image plate detector which was set at a distance of 475 mm from the crystal. One hundred 1° oscillations of the crystal were collected with constant dose at a wavelength of 0.87 Å and an image plate scan diameter of 300 mm. Data processing with MOSFLM (Leslie, 2006), SCALA (Evans, 2006) and other programs in the CCP4 suite (Winn *et al*., 2011) yielded intensity data to a resolution of d_{min} = 2.8 Å. The crystals were found to have 2 ALAD monomers per crystallographic asymmetric unit and a solvent content of 55 %. Other relevant statistics (for all structures) are shown in Table 1. Structure analysis by use of molecular replacement with XPLOR (Brünger *et al*., 1998) and TFFC (Winn *et al*., 2011) using the TIM barrel domain of yeast ALAD (PDBID: 1aw5; Erskine *et al*., 1997a) as the search model was successful and refinement was undertaken by simulated annealing using CNS (Brünger *et al*., 1998) and later REFMAC (Murshudov *et al*., 2011).

X-ray data on recombinant human ALAD were collected at the ESRF (Grenoble) beam line ID14-2 using an ADSC Quantum 4 detector to a resolution of 2.8 Å. A series of 1˚ oscillation images were

recorded with an exposure time of 10 seconds per image and were processed exactly as above. Since, the crystals were not isomorphous with those of the native enzyme, the structure was solved by molecular replacement using the human native ALAD structure as the search model. The program MOLREP (Vagin & Teplyakov, 2010) yielded a solution which was refined, initially with SHELX (Sheldrick & Schneider, 1997), followed by TLS-refinement with RESTRAIN (Haneef *et al*., 1985) and finally REFMAC (Murshudov *et al*., 2011).

X-ray data on Pc-ALAD were collected using station I02 at the Diamond Light Source (DLS, Didcot, England). A series of 1˚ oscillations were used to collect 190˚ of data from a single crystal maintained at a temperature of 100 K using a PILATUS 6M-F detector with an exposure time of 1 second/image (50 % transmission) and a crystal-to-detector distance of 668.8 mm. The incident beam had a wavelength of 1.04346 Å. Data processing was done as above and the structure of an octamer was determined using the molecular replacement program PHASER (McCoy *et al*., 2007) with *E. coli* ALAD (PDB entry 1B4E; Erskine *et al*., 1999) as the search model and was then refined with REFMAC (Murshudov *et al*., 2011). Since the R-free was unsatisfactory at this stage (> 43 %), an effort to improve the model was made by non-crystallographic symmetry (NCS) electron density averaging and solvent flattening with DM (Cowtan, 1994). This yielded an improved map which showed evidence for additional monomers that were not accounted for by the octameric model. Hence MOLREP was run again with a monomeric search model, using the refined octamer as a fixed input structure, and this revealed the locations of another 4 monomers which constitute exactly one half of an additional octamer. Thus the asymmetric unit contains 12 subunits or 1½ octamers, with the other half-octamer being generated by crystallographic symmetry.

Data on the complex of *E. coli* ALAD with PBG were collected at ESRF (Grenoble) on the beam line BM14 with a Marresearch 345 image plate detector using a 30 cm diameter scan. A series of 120 second exposures were collected, each with a rotation angle of 1°, a wavelength of 1.069 Å and a crystal-to-detector distance of 250 mm. Another dataset from a different crystal was collected at DESY (EMBL, Hamburg) using a Marresearch 30 cm image plate detector at the X11 beamline with a wavelength 0.834 Å. The data were collected in two passes: one to a resolution of 2.1 Å (45 two degree oscillations, 180 sec exposures, crystal-to-detector distance 210 mm) and another to measure previously overloaded spots at low resolution (26 three degree oscillations, 30 sec exposures, crystal-to-detector distance 350 mm) both using a 90 mm scan radius. The data were processed as above and since both of the crystals were isomorphous with the native enzyme (Erskine *et al*., 1999), the bound structure of PBG was analysed by calculation of a difference Fourier map following refinement of a water-free model with REFMAC (Murshudov *et al*., 2011). Combining the data from both of these crystals was necessary to give satisfactory outer-shell

completeness. The ligand- and solvent-structure of the complex was built using Coot (Emsley *et al*., 2010) between rounds of refinement.

The final refined structures and reflection datasets have been analysed by the validation programs PROCHECK (Laskowski *et al*., 1993), SFCHECK (Vaguine *et al*., 1999) and MOLPROBITY (Lovell *et al*., 2003; Chen *et al*., 2010) and have been deposited in the Protein Data Bank (www.wwpdb.org) with accession codes that are shown in Table 1. DOIs for the diffraction images are given in the supplementary material. Figures of the structures were prepared using CueMol (www.cuemol.org/en) for which the solvent-accessible surface and electrostatic potentials were calculated using the MSMS (Sanner *et al*., 1996) and APBS (Baker *et al*., 2001) software, respectively. Subunit interface areas were analysed by use of the PISA server (Krissinel & Henrick, 2007).

3. Results

3.1 Human ALAD

There are two monomers in the asymmetric unit of human native ALAD which form a closely associated dimer that, in combination with the crystallographic 4-fold axis, forms octamers with approximate dimensions of 108 Å x 108 Å x 78 Å, in accord with earlier EM studies (Wu *et al*., 1974; Pilz *et al*., 1988). The octamer, the assembly of which is shown in Fig. 2, can be described as a tetramer of dimers with a solvent-filled channel running through the centre. Of the two monomers in the asymmetric unit, monomer A is better defined than the other (monomer B) in the general vicinity of the active site. While both monomers show disorder in the loop region 137 - 138 and at the C-terminal end, monomer B has additional regions of disorder in the loops involving residues 89 - 93, 127 - 142 and 213 - 220, the latter being part of a large loop region which covers the active site of the enzyme. The basic fold of human ALAD is an $(\alpha/\beta)_8$ or TIM barrel that is formed by an 8-membered cylindrical β-sheet surrounded by 8 α-helices. This domain has an N-terminal arm (residues 1 - 28) that engages with the TIM barrel of the adjacent monomer in the dimer. The arm includes a region of distorted 3_{10}-helix (residues 7 - 10) followed by an α-helix (residues 14 - 19) denoted α1. As in all enzymes adopting the $(\alpha/\beta)_8$ fold, the active site is located in an opening formed by the loops connecting the C-terminal ends of the parallel β-strands in the barrel with their ensuing α-helical segments (Cooper & Erskine, 2004). The loop regions between these α and β segments are elaborated extensively at the active site end of the barrel where the regular alternation of helices and strands is broken by the insertion of several extra secondary structure elements. One example of this is the β-hairpin formed which lies between β1 and α2. Here the hydrogen bonds made by the strands of this hairpin (β2 and β3) mean that they are part of the main β-sheet of the molecule and provide an

extension to it at the active site end of the barrel. Another elaboration on the basic TIM-barrel fold is the loop covering the active site (200 - 230) which includes a region of α-helix involving residues 204 - 210 denoted α_{act}. Finally, there is also a helix (α8) that lies between β9 and α9 in the primary sequence and, like other secondary structure elements in the vicinity, is important for quaternary interactions about the non-crystallographic 2-fold axis within the dimer.

The structure of recombinant human ALAD also has two monomers per asymmetric unit which form a symmetric arm-around-barrel hugging dimer, as described above. This dimer superimposes as a rigid group almost exactly with that of the native enzyme – overall 94 % of the amino acids in each dimer superimpose with an RMS deviation of 0.6 Å, which is quite low for the resolutions of these two structures. This suggests that there is little difference in the relative orientations of the constituent monomers within each dimer of the two crystal forms. As for the native structure, the best-defined subunit of the recombinant ALAD dimer has arbitrarily been assigned the letter A and the second subunit, the letter B. There are small differences between the A subunits of the two structures in the region of the following residues: 97 - 98, 128, 137 - 138 (which are missing in the native structure), 214 - 215 and 220. There are more significant differences in the B subunit where residues 83 - 95, 125 - 143 and 213 - 222 are missing in the native structure. These differences between the two structures are all in surface regions and are likely to be due to proteolysis, given that erythrocytes have a lifespan of around 4 months (Shemin & Rittenberg, 1946), disorder in the protein molecule or differences in the crystal contacts.

Superposition of a recombinant human ALAD monomer with the structures of the yeast and *E. coli* ALAD (Fig. 3) gives rms deviations of 0.97 Å (for 311 residues) and 1.48 Å (for 309 residues), respectively. This demonstrates that the *E. coli* enzyme differs more markedly from human ALAD in 3D than the yeast enzyme does and is in accord with the sequence identities - yeast and human ALAD being 53 % identical versus 43 % for the *E. coli* and human sequences. In addition to differences at the extended N-terminus of the protein and its C-terminus, further differences are notable in an exposed loop region involving residues 39 – 44 and a helix (residues 60 - 70) which follows shortly afterwards. Large differences also occur in the long loop region involving residues 84 – 90, in the vicinity of residue 97 and in the loops involving residues 133 – 144, 185 – 190 and 213 – 220. In all of these regions except for the N-terminal tail, which plays an important role in quaternary contacts, the yeast ALAD structure follows that of the human enzyme much more closely than the *E. coli* enzyme does. The human enzyme superimposes with the structure of mouse ALAD (PDBID: 2Z0I) with an RMSD of 0.6 Å for 283 equivalent residues. This very close structural similarity is in accord with the high sequence identity of 89 % for the human and mouse enzymes. Comparison of the

human ALAD structure with that of the most distantly related ALAD, that of *Toxoplasma gondii* which has 40 % sequence identity, reveals an additional region of variability in the range 268 – 270 (human numbering) where the parasite enzyme has a one residue deletion. There is some additional variability in the region of residue 290 and especially at the N- and C-termini where the *T. gondii* enzyme is more extended and has some additional elements of secondary structure which play important roles in octamer formation (Jaffe *et al*., 2011).

The active site of the A subunit of human native ALAD contains a well-defined zinc ion that is coordinated by cysteines 122, 124 and 132 (Fig. 4). The zinc ion lies close to two invariant lysine residues: 199 and 252, which are known from other studies to form Schiff base links with the A- and P-side substrate moieties, respectively. The active site of the A subunit also contains electron density for a ligand molecule which appears to be bound covalently with the active site Lys 252. In accord with previous studies of the yeast enzyme (Erskine *et al*., 1999; 2001b) we have interpreted this as being the substrate ALA bound by a Schiff base link. In contrast, the greater disorder in the active site region of the B subunit renders a number of these features invisible in the electron density map, although Cys 122 and Cys 124 are clearly resolved in positions which would allow them to coordinate a zinc ion. However, residues of the active site loop from 213 to 220 of the native enzyme are disordered and thus missing from the model, as are residues in loops from 90 to 92 and from 127 to 142. The latter region includes the third zinc-binding cysteine (Cys 132) and accordingly the zinc ion was not modelled in this subunit. The structure of the recombinant human enzyme exhibits similar disorder effects in the B subunit although all of these regions have been built into the model and the zinc ion along with its associated cysteines appear to well-defined by the electron density (Fig. 4). The differences between the A subunits of the native and recombinant ALADs are appreciably smaller, being restricted to short regions which are missing from the native enzyme between residues 137 - 138 and 329 - 330 at the C-terminus, as well as local conformational differences in the following regions: 97 - 98 and 214 - 220.

3.2 Comparison of octameric and hexameric human ALAD structures

The structure of a hexameric mutant of human ALAD (F12L) (Breinig *et al*., 2003) showed that the constituent dimers are each twisted by approximately 60° with respect to those of the octameric form. In all ALAD structures the N-terminal arm of one monomer ‘caps-off’ the buried end of the TIM barrel of one of the monomers in an adjacent dimer. In octameric ALADs, this N-terminal arm also ‘hugs’ the TIM barrel of the neighbouring monomer in the same dimer. In contrast, in the hexameric mutant the dimers are oriented such that the N-terminal arm is unable to curl around the TIM barrel of the neighbouring monomer in the dimer, although the barrel-capping interaction with

the adjacent dimer is preserved. Thus superposition of the octamers of all ALAD structures from different species was undertaken to investigate whether there are any relative internal movements of the subunits, due to sequence differences and crystal packing effects, which might indicate how a transition between the octameric and hexameric forms can be made. This was undertaken by using the PISA server (Krissinel & Henrick, 2007) to generate the corresponding octamers, followed by approximate least-squares superposition of the different ALAD octamers using just their P-site lysine residues for simplicity. Viewing these as an animation, one can discern small rotations of the TIM barrel domains around the axes of the barrels as well as small changes in the orientations of the barrel-axes. However these movements do not greatly affect the orientations of the arm-hugging dimers in the octamers and whilst indicating the general range of motion available to the TIM barrel domain of each monomer, unfortunately, they do not give an indication of how the octamer-hexamer transition is made.

The electrostatic surface of the human ALAD octamer is generally dominated by basic and neutral residues (Fig. 5). However, there exists a highly acidic channel running along the four-fold axis of the assembly and this feature appears to be conserved across the enzyme family. The negative charge of this channel is consistent with the fact that a metal-binding site that occurs in ALADs from some species is located within this channel (Erskine *et al*., 1999; Frankenberg *et al*., 1999). The diameter of the channel is in the range of 15 – 20 Å which is comparable to the pore-size of transport proteins and is suggestive of another biological function which ALAD might have, particularly in view of its association with the proteasome and the chaperone Hsp70. When the electrostatic surface of the octamer is viewed from the side with the 4-fold axis vertical, there are four pronounced basic grooves running diagonally across each face of the octamer. Each groove connects the two outward-pointing active sites on each side of the octamer and appears to lead directly into the buried active site cavity of each of the two monomers. The active site itself also has a marked basic character. Thus, if the substrate ALA is negatively charged, these substantial electropositive grooves, each of which has openings into the two active sites of each dimer, could assist catalysis by electrostatic funnelling. At physiological intracellular pH values (around 7.3), the charge of the substrate will be dominated by that of its carboxylate group, which has a pKa of 4.0, rather than by the amino group, which has a pKa of 7.8 and therefore is likely to be only partially positively charged. This suggests that electrostatic funnelling is likely to be a feature of ALAD catalysis and is further corroborated by the fact that the main features of the surface potential of human ALAD are conserved across the range of octamers from species for which ALAD structures have been determined. Alternatively these channels might conceivably provide sites where the product porphobilinogen (which possesses two carboxylate groups) could accumulate prior to

transfer to the next enzyme in the pathway, namely porphobilinogen deaminase. It is also interesting that the electrostatic surface of the inactive hexameric F12L mutant of the human enzyme is appreciably different due to it lacking the central acidic pore and the outer basic grooves. Since the active sites are more central in the hexameric form, the inner regions of this assembly have an abundance of positive charge, although it has to be noted that many residues in loops at the active site end of the TIM barrel are absent in the structure and thus were not included in the electrostatic calculations.

3.3 *P. calidifontis* ALAD

Pc-ALAD expressed in *E. coli* was shown by mass spectrometry to have a molecular mass of 37,676.2 ± 0.2 Da protein. Characterization of Pc-ALAD revealed that it functions optimally at a relatively basic pH range of 7.5 – 9.0 and requires exogenous thiols and zinc ions for optimal activity. Whilst other characterized ALADs can withstand a heat shock of 60 – 70 ˚C for short time period (Jordan and Seehra, 1986), Pc-ALAD was found to be thermostable with a half life of approximately 9 hours at a temperature of 100 ˚C. Analysis of a well-dialyzed Pc-ALAD sample, where all the loosely bound zinc ions had been removed, gave a stoichiometry of 0.7 ± 0.05 mol of zinc per mol of a Pc-ALAD subunit.

Analysis of the X-ray diffraction obtained from Pc-ALAD crystals revealed that they are trigonal and belong to the space group $P3_121$ with unit cell dimensions of $a = b = 205.6$ Å and $c = 199.2$ Å. Molecular replacement confirmed that it is an octameric protein that has crystallised with 12 subunits or 1½ octamers in the asymmetric unit, with the other half-octamer arising from crystallographic symmetry (Fig. 6). This agrees with the reports for other bacterial and animal ALADs which have been found to be octameric (Bevan *et al*., 1980), although recent work has suggested the existence of a less active hexameric form in humans (Sawada *et al*., 2011; Selwood *et al*., 2008) and there are numerous reports that the enzyme is hexameric in higher plants (reviewed in Boese *et al*., 1991).

In spite of the fact that the diffraction data for Pc-ALAD only extend to medium resolution (3.5 Å) the electron density map is of reasonably good quality throughout the structure. This may stem partly from the beneficial effects of the high non-crystallographic symmetry, which was exploited in density averaging and refinement. The side chains of the conserved active site lysines (204 and 257) have good electron density in 10 of the 12 molecules in the asymmetric unit and the other features of the catalytic site, including the bound zinc ion and its coordinating cysteines, are very well-defined in all 12 of the subunits in the asymmetric unit. However, there appears to be a lack of

convincing electron density for the inhibitor, laevulinic acid, which was present during crystallisation. In contrast, there is good electron density in all subunits for a zinc ion bound at the site at which magnesium ions are known to have an allosteric activating effect on the enzyme from some species. Here the metal ion interacts predominantly with the side chains of Glu 242 and Asp 246 although the side chains of Asp 178 and Glu 249 are close by.

An analysis of the subunit interfaces in the Pc-ALAD structure using PISA (Krissinel & Henrick, 2007) indicates that formation of the octamer involves burying around 46,510 $Å^2$ of surface accessible area. This is comparable to the values for ALADs from mesophilic organisms, such as yeast (51,955 $Å^2$) and *E. coli* (41,860 $Å^2$). Since the subunits of these three enzymes are also held together by a similar number of salt-bridges and hydrogen bonds, the basis for the extreme thermostability of the *P. calidifontis* enzyme is not clear, as is often the case in such comparisons. One slightly unusual feature of the Pc-ALAD is that the N-terminal region appears to be appreciably buried within the inter-subunit channel which runs along the central 4-fold axis of the octamer. The equivalent region of other ALAD structures tends to be shorter and / or more solvent-exposed. The fact that this part of four monomers comes close together in a position which essentially seals-off the water-filled channel passing through the centre of the octamer is almost redolent of the safety valve on a boiler! Whether the N-terminal region of Pc-ALAD does contribute specifically to the thermostability of the molecule could, in principle, be tested by deletion mutagenesis. It is interesting that the enzyme from *T. gondii* has a more extensive elaboration at the N-terminal end which forms a β-hairpin turn (Jaffe *et al.*, 2011). However, this hairpin is considerably more exposed than the N-terminal β-strand of Pc-ALAD since it appears to loop-out of the mouth of the channel passing through the octamer, rather than to fold into it. Curiously, the C-terminal end of the *toxoplasma* enzyme also has a β-strand extension which probably plays a role in stability of the octamer by forming an anti-parallel interaction with the same strand of a neighbouring subunit.

3.4 Structure of the non-covalent complex with the product, porphobilinogen

The complex of *E. coli* ALAD with the reaction product PBG has a number of interesting features and differs significantly from a previously reported structure of yeast ALAD with a bound PBG moiety that was covalently attached to the P-site lysine of the enzyme (Erskine *et al.*, 2003). The complex with the yeast enzyme had been obtained by co-crystallisation with the substrate, ALA, rather than by co-crystallisation with the final product, PBG. Thus the covalent bond observed in the yeast ALAD-PBG complex is likely to be a residual effect of the Schiff base link between the P-site lysine residue and the ALA substrate. In contrast, the structure of *E. coli* ALAD possesses a

non-covalently bound PBG moiety which makes very similar hydrogen-bonding and other interactions with the active site residues. The most notable interaction is a dative bond between the pyrrole side chain amino group and the active site zinc ion. This interaction is consistent with previous NMR studies which demonstrated the role of the zinc ion in binding the A-side substrate and the reaction product (Jaffe & Hanes, 1986; Jaffe & Markham, 1987; 1988; Jaffe *et al*., 1990). Curiously, in our structure the pyrrole ring appears to be very elegantly sandwiched between the side chain amino groups of both active site lysine residues (Lys 247 and Lys 195) on one side, and the aromatic ring of Phe 204 which resides in the short region of helix covering the active site, α_{act}, on the other (Fig. 7). Although there are many aromatic side chains in the vicinity of the active site, only Phe 79 and Phe 204 make hydrophobic van der Waals contact with the pyrrole ring atoms. The P-side carboxyl-group of PBG forms hydrogen-bonds with the side chains of Ser 273 and Tyr 312 and the pyrrole ring amino group is devoid of hydrogen bond interactions with the enzyme. This contrasts with the partially formed PBG moiety bound to yeast ALAD in which the ring amino group appears to form hydrogen bonds with conserved serine and aspartate residues. The equivalent residues in *E. coli* ALAD (Asp 118 and Ser 165) are approximately 4 Å from the pyrrole nitrogen in the non-covalent PBG complex and this is too far for hydrogen bonding. Superposition of the two complexes demonstrates that the pyrrole nitrogen is displaced by approximately 1 Å in the two complexes. The A-side carboxyl group is within hydrogen bonding distance of two invariant arginine side chains (Arg 205 and Arg 216) and the side chain of Gln 220 suggesting that this carboxyl side group of the pyrrole is stabilised by both polar and ionic interactions with the enzyme.

4. Discussion

The structures of human ALAD determined from native and recombinant protein are, as expected, substantially similar, although the recombinant protein appears to be better defined than the native enzyme in exposed loop regions. This presumably correlates with the age of protein purified from mature human erythrocytes. These cells are anucleated and therefore do not produce new protein molecules during their lifespan, which can be up to 120 days. Erythrocyte proteins may therefore experience greater degradation during this period than those expressed heterologously in *E. coli* cells, which are usually grown for a matter of hours and the protein is then purified within a period of hours or days while stored at low temperature. The ALADs we have analysed are all zinc-dependent due to the presence of three cysteine residues which coordinate a zinc ion at the active centre. Our analysis of the complex of *E. coli* ALAD with a product molecule non-covalently bound to it demonstrates that the side chain amino group of PBG can occupy the fourth ligand position of the active site zinc ion, suggesting that the amino group of A-side ALA will initially be coordinated in the same way.

The structures we have analysed are all octameric and possess a substantial solvent-filled channel running along the four-fold axis of each octamer. These channels are largely electronegative and accordingly a number of ALADs bind an additional metal ion, usually magnesium, in the channel and this has an activating effect on catalysis. In contrast, the outward-looking faces of the octamer have large electropositive grooves in them which lead into the active sites and these may have a role in funnelling substrate to the catalytic centre.

The ALAD from the hyperthermophile *P. calidifontis* has an unusual feature at its N-terminal end which forms a short region of β-strand that is substantially more buried in the channel running along the four-fold axis of the octamer. This is redolent of the additional secondary structure elements which were found in this region of the *Toxoplasma* enzyme (Jaffe *et al*., 2011). Whilst our analysis of the *P. calidifontis* enzyme does not immediately reveal the basis of its extreme thermostability, it does provide a structural starting-point for theoretical studies of this effect. For example the use of simulation methods for analysing dynamic flexibility and static rigidity have shed much light on the basis of the stability of proteins from thermophilic and psychrophilic organisms (Wells *et al*., 2014).

Acknowledgements

Parts of this work have been funded in the past by the UK Biotechnology and Biological Sciences Research Council (BBSRC). We gratefully acknowledge the UK Science and Technology Facilities Council (STFC), the European Molecular Biology Laboratory at the Deutsches Elektronen-Synchrotron (DESY, Hamburg) and the European Synchrotron Radiation Facility (ESRF, Grenoble) for synchrotron beam time and associated user support. We are very grateful to the anonymous referee who contributed some re-written paragraphs to the revised manuscript.

References

Ajioka, R. S., Phillips, J. D. & Kushner, J. P. (2006). *BBA* **1763**, 723-736.

Amo, T., Paje, M. L., Inagaki, A, Ezaki, S., Atomi, H. & Imanaka, T. (2002). *Archaea*. **1**, 113-121.

Baker, N. A., Sept, D., Simpson, J., Holst, M. J. & McCammon, J. A. (2001). *Proc. Natl. Acad. Sci. USA* **98**, 10037-10041.

Bardag-Gorse, F. & French, S. W. (2011). *Exp. Molec. Pathol*. **91**, 485-489.

Bevan, D. R., Bodlaender, P. & Shemin, D. (1980). *J. Biol. Chem*. **255**, 2030-2035.

Boese, Q.F., Spano, A. J., Li, J. & Timko, M. P. (1991). *J. Biol. Chem*. **266**, 17060-17066.

Breinig, S., Kervinen, J., Stith, L., Wasson, A. S., Fairman, R., Wlodawer, A., Zdanov, A. & Jaffe, E. K. (2003). *Nat. Struct. Biol*. **10**, 757-763.

Brünger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J.-S., Kuszewski, J., Nilges, N., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. & Warren, G. L. (1998). *Acta Crystallogr. D* **54**, 905–921.

Chen, V. B., Arendall, W. B., Headd, J. J., Keedy, D. A., Immormino, R. M., Kapral, G. J., Murray, L. W., Richardson, J. S. & Richardson, D. C. (2010). *Acta Crystallogr. D* **66**, 12-21.

Coates, L., Beaven, G., Erskine, P. T., Beale, S. I., Avissar, Y. J., Gill, R., Mohammed, F., Wood, S. P., Shoolingin-Jordan, P. & Cooper, J. B. (2004). *J. Molec. Biol*. **342**, 563-570.

Cooper, J. B. & Erskine, P. T. (2004). In 'Handbook of Metalloproteins. Vol. 3' Ed. Messerschmidt, A., Bode, W. & Cygler, M. Wiley, Chichester, pp. 283-295.

Cowtan, K. (1994). *Joint CCP4 and ESF-EACBM Newsletter on Protein Crystallography*. **31**, 34-38.

Doss, M., Von-Tieperman, R., Schneider, J. and Schmid, H. (1979). *Klin. Wochenschr*. **57**, 1123-1127.

Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. (2010). *Acta Crystallogr D* **66**, 486-501.

Erdtman, E., Bushnell, E. A., Gauld, J. W. & Eriksson, L. A. (2010). *J. Phys. Chem. B*. **114**, 16860-16870.

Erskine, P. T., Norton, E., Cooper, J. B., Lambert, R., Coker, A., Lewis, G., Spencer, P., Sarwar, M., Wood, S. P., Warren, M. J. & Shoolingin-Jordan, P. M. (1999). *Biochemistry* **38**, 4266-4276.

Erskine, P. T., Senior, N., Awan, S., Lambert, R., Lewis, G., Tickle, I. J., Sarwar, M., Spencer, P., Thomas, P., Warren, M. J., Shoolingin-Jordan, P. M., Wood, S. P. & Cooper, J. B. (1997a). *Nat. Struct. Biol*. **4**, 1025-1031.

Erskine, P. T., Senior, N., Maignan, S., Cooper, J., Lambert, R., Lewis, G., Spencer, P., Awan, S., Warren, M., Tickle, I. J., Thomas, P., Wood, S. P. & Shoolingin-Jordan, P. M. (1997b). *Protein Sci*. **6**, 1774-1776.

Erskine, P. T., Coates, L., Butler, D., Youell, J. H., Brindley, A. A., Wood, S. P., Warren, M. J., Shoolingin-Jordan, P. M. & Cooper, J. B. (2003). *Biochem. J*. **373**, 733-738.

Erskine, P. T., Coates, L., Newbold, R., Brindley, A. A., Stauffer, F., Wood, S. P., Warren, M. J., Cooper, J. B., Shoolingin-Jordan, P. M. & Neier, R. (2001a). *FEBS Lett*. **503**, 196-200.

Erskine, P. T., Newbold, R., Brindley, A. A., Wood, S. P., Shoolingin-Jordan, P. M., Warren, M. J. & Cooper, J. B. (2001b). *J. Molec. Biol*. **312**, 133-141.

Evans, P. (2006). *Acta Crystallogr*. *D* **62**, 72-82.

Frankenberg, N., Erskine, P. T., Cooper, J. B., Shoolingin-Jordan, P. M., Jahn, D. & Heinz. D. W. (1999). *J. Mol. Biol*. **289**, 591-602.

Gibbs, P. N. B. & Jordan, P. M. (1986). *Biochem. J*. **236**, 447-451.

Gibbs, P. N., Chaudhry, A. G. & Jordan, P. M. (1985). *Biochem. J*. **230**, 25-34.

Guo, G. G., Gu, M. & Etlinger, J. D. (1994). *J. Biol. Chem*. **269**, 12399-12402.

Gross, M., Hessefort, S. & Olin, A. (1999). *J. Biol. Chem*. **274**, 3125-3134.

Haneef, I., Moss, D. S., Stanford M. J. & Borkakoti, N. (1985). *Acta Crystallogr*. *A* **41**, 426–33.

Heinemann, I. U., Schulz, C., Schubert, W.-D., Heinz, D. W., Wang, Y.-G., Kobayashi, Y., Awa, Y., Wachi, M., Jahn, D. & Jahn, M. (2010). *Antimicro. Agents Chemo*. **54**, 267-272.

Holton, T. A. & Graham, M. W. (1991). *Nucl. Acid. Res*. **19**, 1156.

Jaffe, E. K. (1995). *J. Bioenerg. Biomembr*. **27**, 169-179

Jaffe, E. K. (2003). *Chem. Biol*. **10**, 25-34.

Jaffe, E. K. (2004). *Bioorg Chem*. **32**, 316-325.

Jaffe, E. K. (2016). *Acc. Chem. Res*. **49**, 2509-2517.

Jaffe, E. K., Ali, S., Mitchell, L. W., Taylor, K. M., Volin, M. & Markham, G. D. (1995). *Biochemistry* **34**, 244-251.

Jaffe, E. K. & Hanes, D. (1986). *J. Biol. Chem*. **261**, 9348-9353.

Jaffe, E. K. & Lawrence, S. H. (2012). *Arch. Biochem. Biophys*. **519**, 144-153.

Jaffe, E. K. & Lawrence, S. H. (2014). *Handbook of Porphyrin Science*, edited by C. G. Ferreira Vol 26, pp 79-128. Singapore: World Scientific.

Jaffe, E. K. & Markham, G. D. (1987). *Biochemistry* **26**, 4258-4264.

Jaffe, E. K. & Markham, G. D. (1988). *Biochemistry* **27**, 4475-4481.

Jaffe, E. K., Markham, G. D. & Rajagopalan, J. S. (1990). *Biochemistry* **29**, 8345-8350.

Jaffe, E. K., Martins, J., Li, J., Kervinen, J. & Dunbrack, R. L. (2001). *J. Biol. Chem*. **276**, 1531-

1537.

Jaffe, E. K., Shanmugan, D., Gardberg, A., Dieterich, M., Banumathi, S., Stewart, L. J., Myler, P. J. & Roos, D. S. (2011). *J. Biol. Chem*. **286**, 15298-15307.

Jaffe, E. K. & Stith, L. (2007). *Am. J. Hum. Genet*. **80**, 329-337.

Jordan, P. M. & Gibbs, P. N. (1985). *Biochem. J*. **227**, 1015–1020.

Jordan, P. M. & Seehra, J. S. (1986). *Methods Enzymol*. **123**, 427-434.

Jordan, P. M. (1991). *New Comprehensive Biochemistry* **19**, 1-66.

Jordan, P. M. (1994). *Curr. Opin. Struc. Mol. Biol*. **4**, 902-911.

Kantardjieff, K. A. and Rupp, B. (2003). *Prot. Sci.* **12**, 1865-1871.

Karplus, P. A. & Diederichs, K. (2012). *Science* **336**, 1030-1033.

Kaya, A. H., Plewinska, M., Wong, D. M., Desnick, R. J. & Wetmur, J. G. (1994). *Genomics* **19**, 242-248.

Kokona, B., Rigotti, D. J., Wasson, A. S., Lawrence, S. H., Jaffe, E. K. & Fairman, R. (2008). *Biochemistry* **47**, 10649-10656.

Krissinel, E. & Henrick, K. (2007). *J. Molec. Biol*. **372**, 774-797.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Crystallogr*. **26**, 283-291.

Layer, G., Reichelt, J., Jahn, D. & Heinz, D.W. (2010). *Prot. Sci*. **19**, 1137-1161.

Lentz, C. S., Halls, V. S., Hannam, J. S., Strassel, S., Lawrence, S. H., Jaffe, E. K., Famulok, M., Hoerauf, A. & Pfarr, K. M. (2014). *J. Med. Chem*. **57**, 2498-2510.

Lentz, C. S., Stumpfe, D., Bajorath, J., Famulok, M., Hoerauf, A. & Pfarr, K. M. (2013). *Bioorg. Med. Chem. Lett*. **23**, 5558-5562.

Leslie, A. G. W. (2006). *Acta Crystallogr. D* **62**, 48-57.

Lindblad, B., Lindstedt, S. & Steen, G. (1977). *Proc. Natl. Acad. Sci. USA* **74**, 4641-4645.

Lovell, S. C., Davis, I. W., Arendall, B., de Bakker, P. I. W., Word, J. M., Prisant, M. G., Richardson, J. S. & Richardson, D. C. (2003). *Proteins: Struct. Func. Genet*. **50**, 437-450.

Mauzerall, D. & Granick, S. (1956). *J. Biol. Chem*. **219**, 435-446.

McCoy, A. J., Grosse-Kunstleve, R. W., Adams, P. D., Winn, M. D., Storoni, L. C. & Read, R. J. (2007). *J. Appl. Cryst*. **40**, 658-674.

Murshudov, G. N., Skubák, P., Lebedev, A. A., Pannu, N. S., Steiner, R. A., Nicholls, R. A., Winn, M. D., Long, F. & Vagin, A. A. (2011). *Acta Crystallogr. D* **67**, 355-367.

Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). *Acta Crystallogr. D* **53**, 240-255.

Pilz, I., Schwarz. E., Vuga, M. & Beyersmann, D. (1988). *Biol. Chem. Hoppe-Seyler* **369**, 1099–1104.

Potluri, V. R., Astrin, K. H., Wetmur, J. G., Bishop, D. F. & Desnick, R. J. (1987). *Hum. Genet*. **76**, 236-239.

Powell, H. R., Johnson, O. & Leslie, A. G. W. (2013). *Acta Crystallogr. D* **69**, 1195-1203.

Read, R. J. (1986). *Acta Crystallogr. A* **42**, 140-149.

Sanner, M. F., Spehner, J. - C. & Olson, A. J. (1996). *Biopolymers* **38**, 305-320.

Sassa, S. (1998). *Sem. Liver Dis*. **18**, 95-101.

Sawada, N., Nagahara, N., Arisaka, F., Mitsuoka, K. & Minami, M. (2011). *Amino Acids* **41**, 173-180.

Schuettelkopf, A. W. & van Aalten, D. M. F. (2004). *Acta Crystallogr. D* **60**, 1355-1363.

Selwood, T., Tang, L., Lawrence, S.H., Anokhina, Y. & Jaffe, E.K. (2008). *Biochemistry* **47**, 3245-3257.

Sheldrick, G. M. & Schneider, T. R. (1997). *Methods Enzymol*. **277**, 319-343.

Shemin, D. & Rittenberg, D. (1946). *J. Biol. Chem*. **166**, 627-636.

Simons, T. J. B. (1995). *Eur. J. Biochem*. **234**, 178-183.

Spencer, P. & Jordan, P. (1993). *Biochem. J*. **290**, 279-287.

Spencer, P. & Jordan, P. M. (1994) *Ciba Found. Symp*. **180**, 50-69.

Tian, B. X., Erdtman, E. & Eriksson, L. A. (2012). *J. Phys. Chem. B*. **116**, 12105-12112.

Vagin, A. & Teplyakov, A. (2010). *Acta Crystallogr. D* **66**, 22-25.

Vaguine, A. A., Richelle, J. & Wodak, S. J. (1999). *Acta Crystallogr. D* **55**, 191-205.

Warren, M. J. & Smith, A. G. (2009). Tetrapyrroles: Birth, Life and Death. Landes Biosciences, Austin, Texas, USA.

Warren, M. J. and Scott, A. I. (1990). *Trends Biochem. Sci*. **15**, 486-491.

Warren, M. J., Cooper, J. B., Wood, S. P. and Shoolingin-Jordan, P. M. (1998). *Trends Biochem. Sci*. 23, 217-221.

Wells, S. A., Crennell, S. J. & Danson, M. (2014). *Prot. Struc. Func. Bioinf*. **82**, 2657-2670.

Wetmur, J. G., Bishop, D. F., Cantelmo, C. & Desnick, R. J. (1986). *Proc. Nat. Acad. Sci*. **83**, 7703-7707.

Winn, M. D. *et al*., (2011). *Acta Crystallogr. D* **67**, 235-242.

Wu, W. H., Shemin, D., Richards, K. E. & Williams, R. C. (1974). *Proc. Natl. Acad. Sci. USA* **71**, 1767–1770.

Table 1. X-ray statistics for the ALAD structures. Values in parenthesis pertain to the outer resolution shell.

ALAD	Human native	Human recombinant	*P. calidifontis*	*E. coli* PBG complex
Beamline	SRS 9.5	ID14-2 (ESRF)	Diamond I02	ESRF BM14, DESY X11
Wavelength (Å)	0.87	0.93	1.04	1.07, 0.83
Space group	*I*422	$P42_12$	$P3_121$	*I*422
Unit-cell parameters				
a, *b* (Å)	125.5	127.1	205.6	128.6
c (Å)	200.9	91.2	199.2	142.8
Mosaic spread (°)	0.5	0.7	0.3	0.5
Resolution (Å)	54.2-2.8 (3.0-2.8)	29.2-2.8 (3.0-2.8)	99.6-3.5 (3.6-3.5)	36.8-2.1 (2.2-2.1)
R_{merge}* (%)	7.1 (25.7)	13.4 (39.3)	11.8 (54.8)	13.1 (63.2)
R_{meas}# (%)	8.2 (38.3)	14.2 (41.5)	13.2 (61.1)	13.7 (67.3)
Completeness (%)	96.0 (76.7)	99.7 (98.6)	100.0 (100.0)	99.9 (99.4)
Average $I/\sigma(I)$	6.0 (2.9)	11.1 (5.3)	14.7 (4.8)	15.9 (3.3)
Multiplicity	5.9 (3.7)	9.2 (9.1)	9.8 (9.9)	11.5 (8.4)
N observed reflections	110,047 (11,426)	173,694 (24,357)	622,972 (44,178)	401,332 (42,004)
N unique reflections	18,652 (3,088)	18,874 (2,665)	63,352 (4,448)	34,864 (4,971)
Wilson plot *B* factor (Å^2)	36.1	69.1	75.9	25.3
Solvent content (%)	55.2	51.9	54.2	70.3
Refinement				
R-factor (%)	16.5	17.2	18.1	15.7
Free *R*-factor (%)	23.8	26.5	25.1	21.3
N refl. working set	17,779	17,923	60,148	32,987
N refl. free set	872	973	3,154	1,780
Ramachandran allowed (%)	99	99	97	100
RMSD bond lengths (Å)	0.01	0.01	0.01	0.03
RMSD bond angles (°)	1.61	1.56	1.7	2.5
Mean *B*-factor (Å^2)	57.2	51.4	79.4	35.2
PDB identifier	5hnr	5hms	5lzl	5mhb

* $$R_{\text{merge}} = \frac{\sum_h \sum_j \left| I_{hj} - I_h \right|}{\sum_h \sum_j I_{hj}}$$

\# $$R_{\text{meas}} = \frac{\sum_h \left(\frac{N_h}{N_h - 1} \right)^{1/2} \sum_j \left| I_{hj} - I_h \right|}{\sum_h \sum_j I_{hj}}$$

where I_h is the mean intensity of the N_h observations I_{hj} of each unique reflection h after scaling.

Figure legends

Fig. 1. The reaction catalysed by 5-aminolaevulinic acid dehydratase or porphobilinogen synthase (PBGS). The two 5-aminolaevulinic acid moieties forming the acetate (A) and propionate (P) side groups of the product, porphobilinogen, are shown. The next enzyme in the pathway, porphobilinogen deaminase (PBGD), catalyses the condensation of four porphobilinogen moieties to form the first linear tetrapyrrole of the pathway. The following steps in the pathway lead ultimately to the formation of several key cyclic tetrapyrrole cofactors, including haem and chlorophyll.

Fig. 2. The structure and assembly of human native ALAD. (a) Shows the A-monomer of the enzyme (helices coloured pale red and strands light green) looking along the TIM-barrel axis with the large flap covering active site, facing towards the viewer. The catalytic lysine residues, 199 and 252, are shown in ball-and-stick representation. (b) The asymmetric unit of the structure contains a dimer, the second monomer of which (monomer B) is less well defined and is shown on top in this view (helices cream and strands blue). It can be seen that the large flap covering the active site of monomer B is not defined. (c) Shows the assembly of the complete ALAD octamer with the A- and B-subunits coloured as in (a) and (b). The active site zinc ions which are present in the A-monomers are indicated as grey spheres.

Fig. 3. Superposition of the recombinant human ALAD monomer with yeast and *E. coli* ALAD. The human enzyme is coloured pink, yeast ALAD is coloured green and *E. coli* is shown in yellow. The active site zinc ions are indicated as grey spheres.

Fig. 4. The active site of human native and recombinant ALAD. The electron density contoured at 1.0 rms for the active sites of the A and B subunits of human native ALAD are shown in (a) and (b), respectively. The A subunit has additional feature of electron density for a ligand which we have putatively interpreted as a substrate molecule (ALA) bound by a Schiff base to Lys 252. The active sites of the A and B subunits of the human recombinant enzyme are shown in (c) and (d), respectively, contoured at the same level. The recombinant enzyme has better definition of the zinc-binding cysteines in both subunits but lacks a bound substrate molecule. Note that the dative bonding to the zinc ion (grey) is only drawn in (a) since this structure possesses electron density for the full complement of tetrahedral ligands, the fourth ligand being a water molecule.

Fig. 5. The electrostatic surface of the human recombinant ALAD octamer. (a) Shows a cut-away view looking along the 4-fold axis and emphasises the negative charge (red) associated with the

inner channel. The positively charged grooves associated with the outer surface of the octamer are also apparent (blue). These can be seen to form diagonals connecting the active sites of adjacent subunits, as shown in (b) in which the octamer is viewed from the same direction as Fig. 2(c) with the 4-fold axis is vertical.

Fig.6. Crystal packing of Pc-ALAD. The 12 monomers which constitute the crystallographic asymmetric unit are coloured gold. These form one complete octamer (shown on the left) and another half-octamer (shown on the bottom right). However, the crystallographic symmetry generates a complete octamer from neighbouring monomers (shown in pale blue, top right).

Fig. 7. The structure of PBG bound to *E. coli* ALAD. The refined $2F_o$-F_c electron density for the PBG and the active site zinc ion (grey) is shown in blue lines contoured at 1.0 rms. The main interacting residues, including the Schiff-base lysines in the foreground, are labelled and hydrogen bonds indicated by dotted lines.

Figures

Fig. 1.

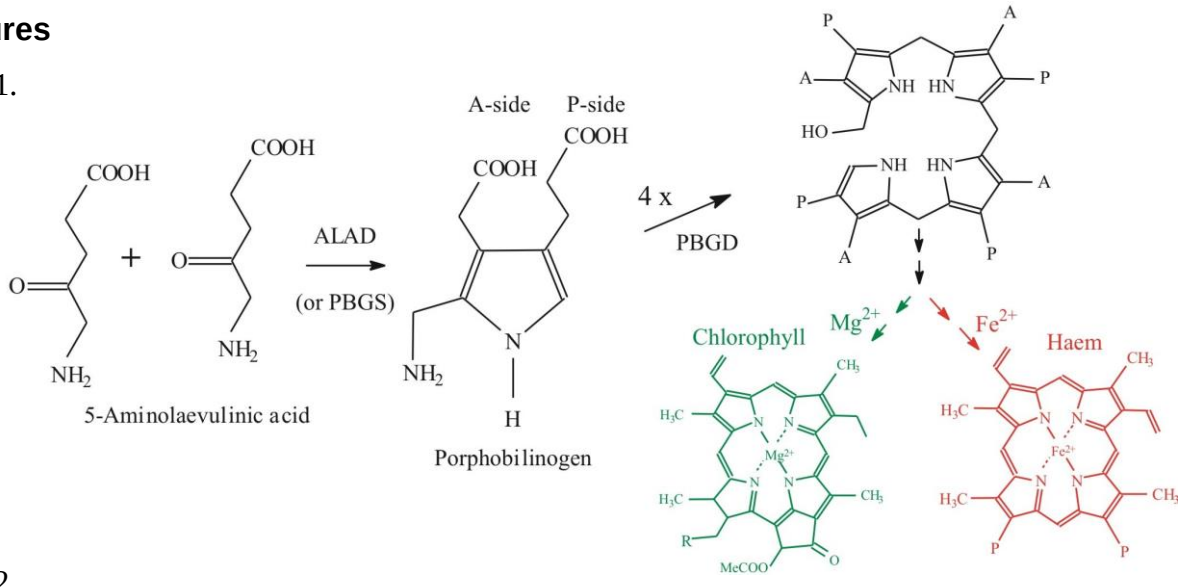


Fig. 2.

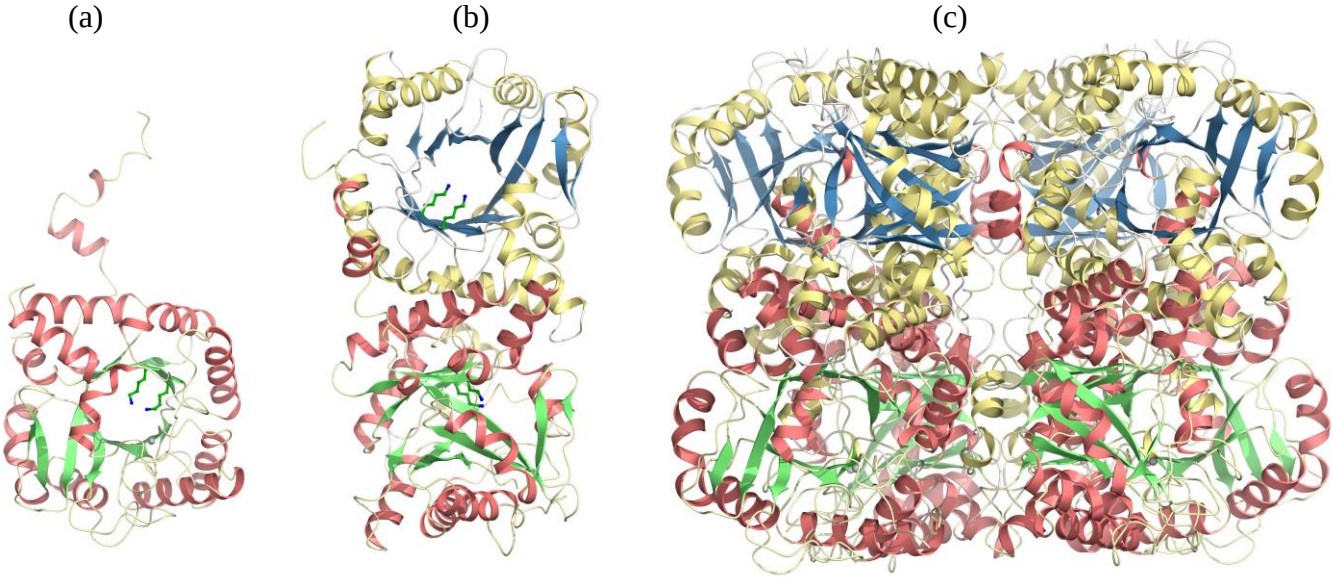
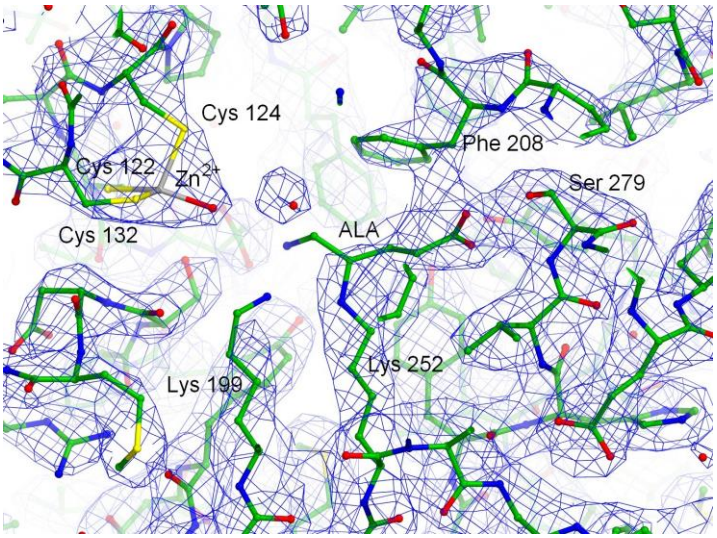


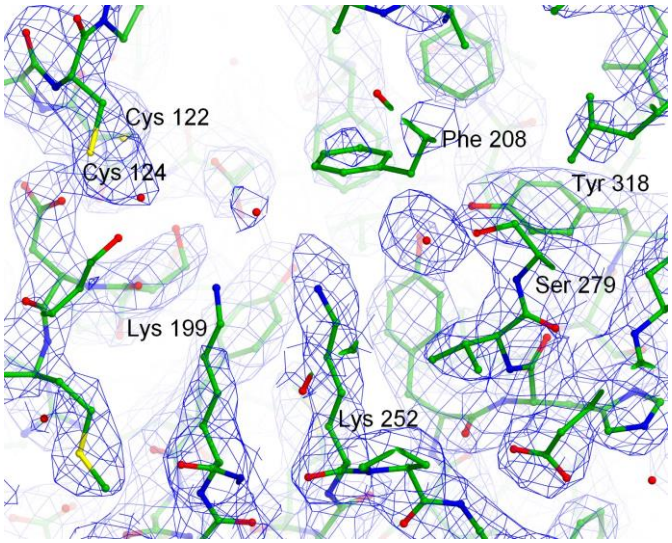
Fig. 3.

Fig. 4.

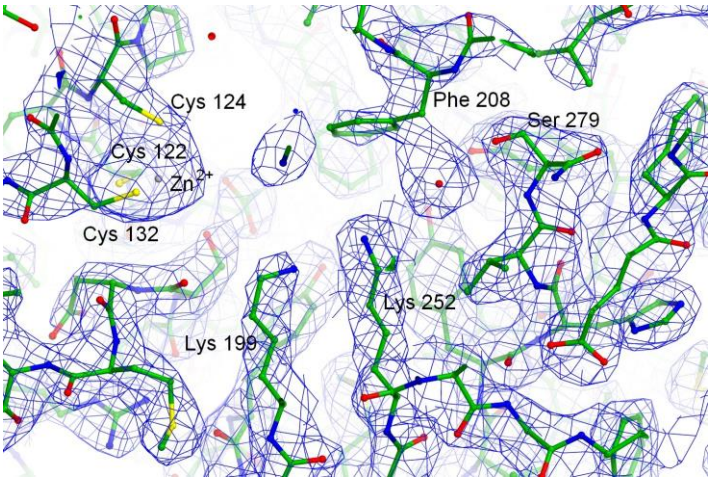
(a)


Cys 124
Phe 208
Cys 122
Zn2+
Ser 279
Cys 132
ALA
Lys 199
Lys 252

(b)


Cys 122
Phe 208
Cys 124
Tyr 318
Ser 279
Lys 199
Lys 252

(c)


Cys 124
Phe 208
Ser 279
Cys 122
Zn2+
Cys 132
Lys 252
Lys 199

(d)

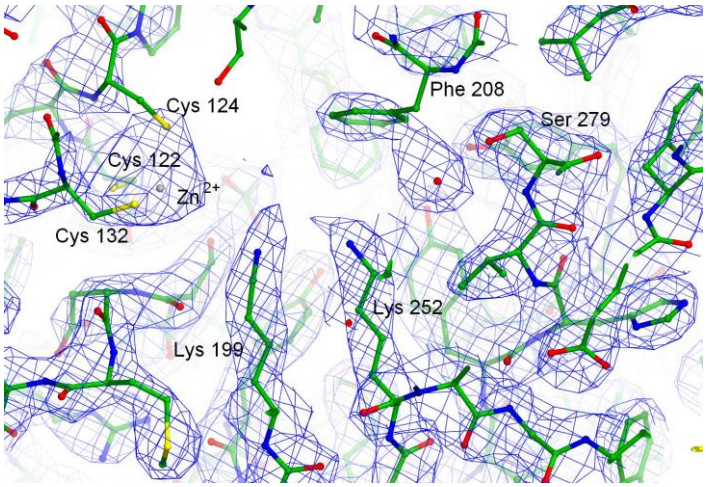

Cys 124
Phe 208
Ser 279
Cys 122
Zn2+
Cys 132
Lys 252
Lys 199

Fig. 5.

(a) (b)

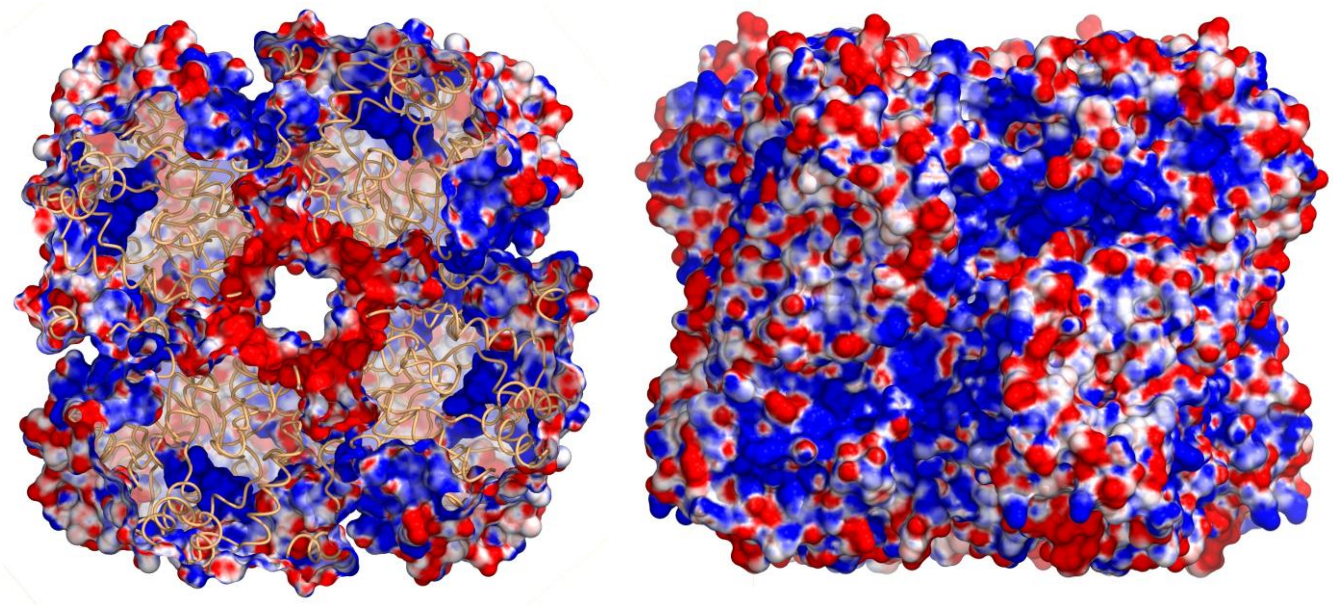


Fig. 6.

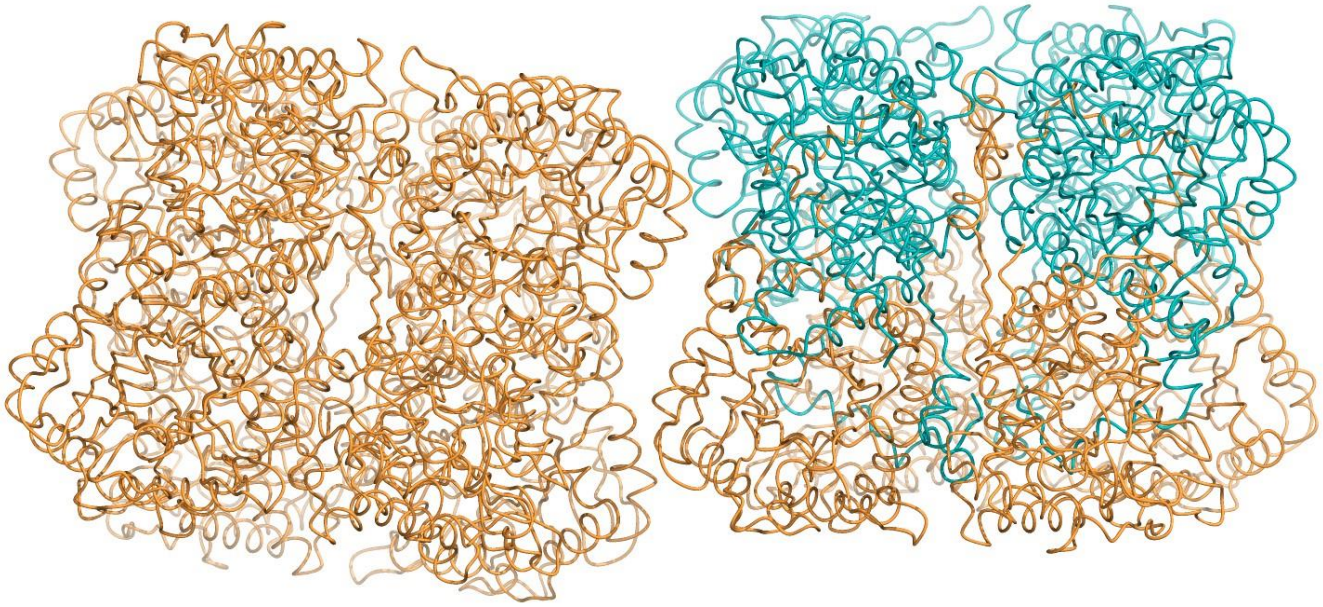
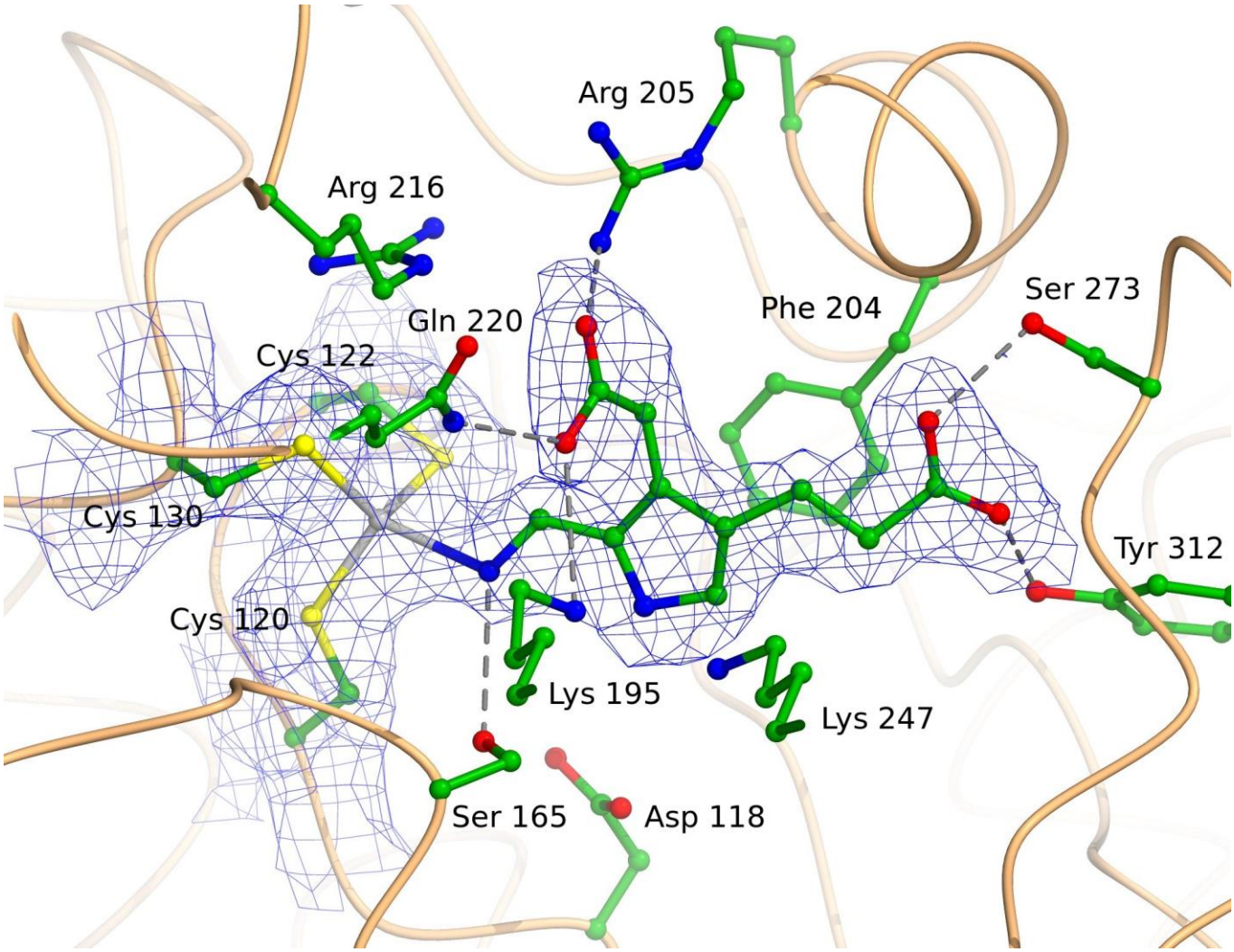


Fig. 7.


Arg 205
Arg 216
Gln 220
Cys 122
Phe 204
Ser 273
Cys 130
Tyr 312
Cys 120
Lys 195
Lys 247
Ser 165
Asp 118